PATENT ABSTRACTS OF JAPAN

(11)Publication number : **07-084592**
(43)Date of publication of application : **31.03.1995**

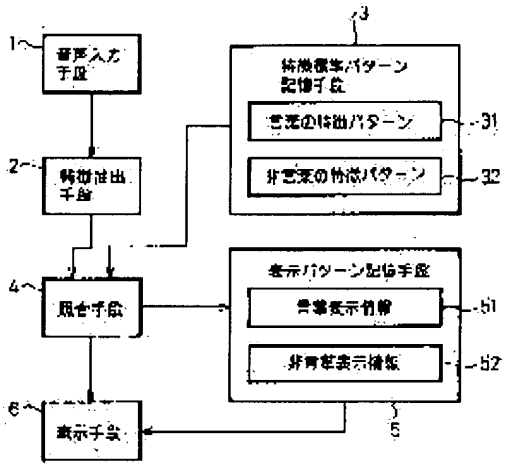
(51)Int.Cl. G10L 3/00
G10L 3/00

(21)Application number : **05-228990**
(22)Date of filing : **14.09.1993**

(71)Applicant : **FUJITSU LTD**
(72)Inventor : **IWAMIDA HITOSHI**

(54) SPEECH RECOGNITION DEVICE

(57)Abstract:
PURPOSE: To recognize a sound other than human words and gives information on it.
CONSTITUTION: The speech recognition device is equipped with a speech input means 1, a feature extracting means 2 which extracts features of a speech signal, a feature standard pattern storage means 3 which stores the feature pattern of the a standard speech signal, a matching means 4 which collates the extracted features of the speech input signal with the stored feature pattern and specifies the standard speech signal corresponding to the input speech signal, a display pattern storage means 5 which stores display information corresponding to the standard speech signal, and a display means 6 which displays the display information; and the feature standard pattern storage means 3 stores the feature patterns 31 of human words and feature patterns 32 other than the human words and the display pattern storage means 5 stores word display information 51 generated by representing the standard speech signal with words as they are and non-word display information 52 corresponding to a non-word standard speech signal.



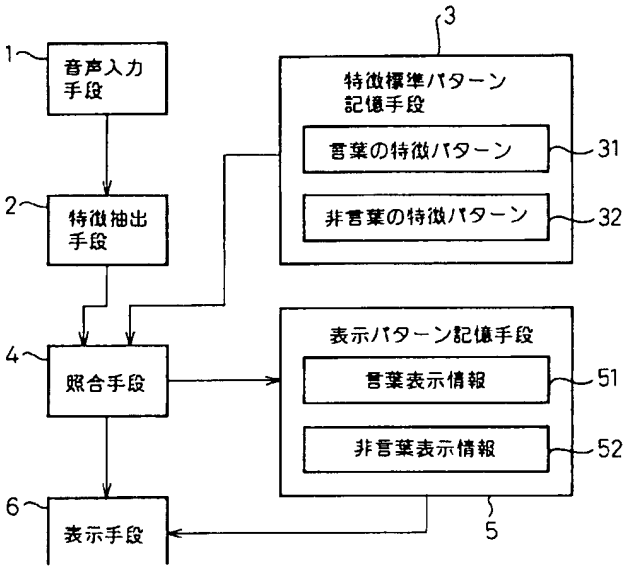
LEGAL STATUS

[Date of request for examination] 19.11.1999
[Date of sending the examiner's decision of rejection] 16.07.2002
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

DRAWINGS

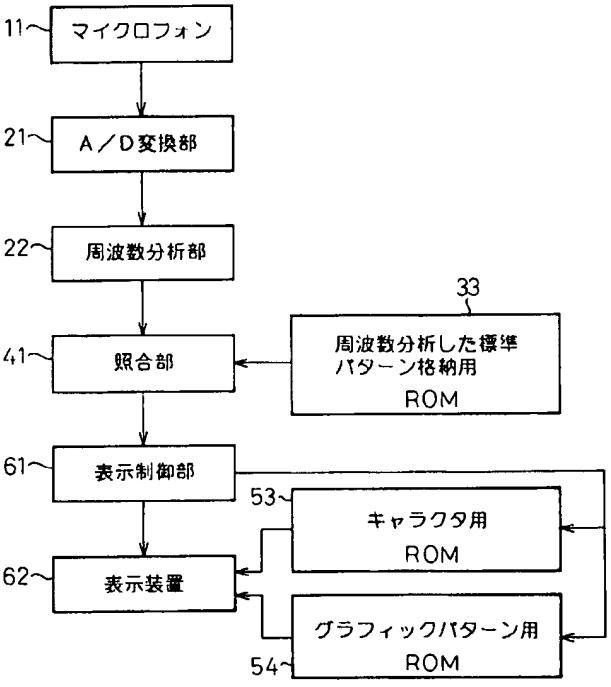
[Drawing 1]

本発明の原理構成図



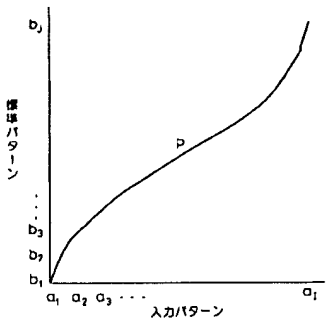
[Drawing 2]

実施例の装置の構成



[Drawing 3]

DPマッチングの説明図



[Drawing 4]
非言葉の表示パターン

(1) 文字表示

(a) サイレンの音

サイレンが鳴っています。

(b) 赤坊の泣き声

赤ちゃんが泣いています。

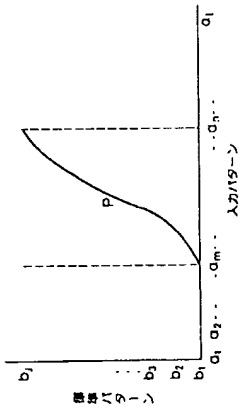
(2) グラフィック表示

(a) サイレンの音

(b) 赤坊の泣き声

[redacted]

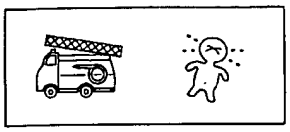
[Drawing 5]
DPマッチングの変形例



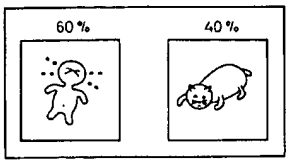
[Drawing 6]

表示の他の例

(1) 同時に2種の音



(2) 類似の音を完全に識別できない時



[Translation done.]

* NOTICES *

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the principle block diagram of the voice recognition unit of this invention.
[Drawing 2] It is drawing showing the composition of the equipment of the example of this invention.
[Drawing 3] It is explanatory drawing of DP matching.
[Drawing 4] It is drawing showing the example of the display pattern of non-language.
[Drawing 5] It is explanatory drawing of the modification of DP matching.
[Drawing 6] It is drawing showing the modification of the display pattern of non-language.
[Description of Notations]
1 -- Voice input means
2 -- Feature-extraction means
3 -- The feature standard-pattern storage means
4 -- Collating means
5 -- Display pattern storage means
6 -- Display means
31 -- The feature pattern of language
32 -- The feature pattern of non-language
51 -- Language display information
52 -- Non-language display information

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] this invention carries out automatic recognition of the inputted sound signal, and relates to the voice recognition unit displayed in the form where sound signals other than human being's words are also recognized, and the sound signal can be recognized especially about the voice recognition unit which displays the recognition result.
[0002]
[Description of the Prior Art] A voice recognition unit is being put in practical use. The conventional voice recognition unit registered the sound signal which should be recognized beforehand as a standard pattern, was memorized, and collated the inputted sound signals and these standard patterns, and the standard pattern which is best in agreement was specified, and when the coincidence condition is more than predetermined level, it is recognized as the voice of a standard pattern having been inputted. Although it is restricted to the sound signal which needs to input the standard pattern and a user can usually input as a standard pattern in order to register as a standard pattern, when recognizing an extraordinary noise for automatic control, such as a plant, such an extraordinary noise may be made into a standard pattern.
[0003] Although a recognition result may be outputted as it is as that to which recognition of a sound signal was performed correctly, the voice which compounded the accuracy of recognition to the term ** sake according to the recognition result is outputted, and, generally it is performed that I have a speaker check. although there is also a method of there being nothing, and displaying a recognition result as alphabetic information and checking it in it as a synthesized speech is outputted to this symptom, make it any -- it is required to be the method of expressing the inputted sound signal correctly Therefore, human being's talk words are made into a standard pattern, and it enables it to perform speech synthesis for a check, or character representation. When the above-mentioned extraordinary noise is made into a standard pattern, since it is not necessary to check, a check is not performed.
[0004] Although it is based on sign language or **** when a hearing-impaired person talks with the others, it is possible to support talking with those to whom a hearing-impaired person has the usual competence using the above-mentioned voice recognition unit. Those who have the usual competence talk by usually passing, recognize the voice with a voice recognition unit, display a recognition result as alphabetic information, and inform a hearing-impaired person of it.
[0005] Although this invention is suitable for the voice recognition unit especially used for such the purpose, it is applicable also to the voice recognition unit which it is in the state which is not desirable for it not to be restricted to such a thing, for example, to emit sound like [under meeting], and is used when the speech information of somewhere else needs to be known. However, since the voice recognition unit for hearing-impaired persons is considered to be the most suitable as an example, this is explained as an example here.
[0006]
[Problem(s) to be Solved by the Invention] In the voice recognition unit which displays the recognized above sound signals, since it was required to be able to display the recognized sound signal, the standard pattern registered was limited to human being's talk words. Moreover, since it was a character corresponding to human being's talk words, what is displayed should just have been what display can also display a character as.
[0007] If it is only for the voice recognition unit for hearing-impaired persons supporting conversation with those who have the usual competence, it will be thought that the voice recognition unit with which it is limited to human being's talk words, and display can also display a character is enough as the above standard patterns registered. However, it is desirable that speech information other than human being's talk words can also be offered from a viewpoint of offering latus speech information by the hearing-impaired person.
[0008] this invention aims at realization of the voice recognition unit which can also offer speech information other than human being's talk words as display information from such a viewpoint.
[0009]
[Means for Solving the Problem] Drawing 1 is the principle block diagram of the voice recognition unit of this invention. A voice input means 1 by which the voice recognition unit of this invention inputs a sound signal like illustration, A feature-extraction means 2 to extract the feature for recognizing a sound signal, and a feature standard-pattern storage means 3 to memorize the feature pattern of a standard sound signal, A collating means 4 to specify the standard sound signal corresponding to the sound signal which collated the feature of the extracted voice input signal, and the feature pattern memorized by the feature standard-pattern storage means 3, and was inputted, In a voice recognition unit equipped with a

display pattern storage means 5 to memorize the display information corresponding to a standard sound signal, and a display means 6 to display the display information corresponding to a standard sound signal when a standard sound signal is specified with the collating means 4 In order to attain the above-mentioned purpose, the feature standard-pattern storage means 3 The feature pattern 32 of non-language standard sound signals other than human being's words other than the feature pattern 31 of the standard sound signal human being's words is memorized. the display pattern storage means 5 It is characterized by having memorized the language display information 51 which made the standard sound signal the character as it was, and the non-language display information 52 corresponding to a non-language standard sound signal.

[0010]

[Function] In the voice recognition unit of this invention, if sound signals other than human being's words, for example, the sound of a "siren", are memorized as a feature pattern 32 of a non-language standard sound signal for the feature standard-pattern storage means 3, a siren's sound can be recognized. And when a siren's sound has been recognized, the display memorized by the display pattern storage means 5 as non-language display information 52 corresponding to a siren's sound "the siren is sounding" can be performed.

[0011] Moreover, if it is the sound of the "siren" of a motor fire engine and the picture of a motor fire engine will be displayed since it cannot distinguish whether the others have said, "The siren is sounding" and whether the sound of a "siren" can actually be heard only by saying ["the siren is sounding"] for example, it will become possible to offer bigger information.

[0012]

[Example] Drawing 2 is drawing showing the composition of the voice recognition unit of one example of this invention, and is a voice recognition unit for a hearing-impaired person. In drawing 2 , the A/D converter which the microphone which changes a sound signal into an electrical signal, and 21 sample the electrical signal acquired from a microphone 11 by the 12kHz sampling period, and changes a reference number 11 into a digital signal, and 22 are the frequency-analysis sections. The frequency-analysis section 22 carries out frequency analysis of the digital time series signal by which A/D conversion was carried out by the first Fourier transform (FFT) etc. every 10ms, asks for the power in each frequency band divided into eight bands with the acoustic-sense-scale, and obtains the time series of a frequency feature parameter. 33 is ROM for standard-pattern storing, and memorizes the time series of the frequency feature parameter of the signal for recognition beforehand searched for in a microphone 11, above-mentioned A/D converter 21, and the above-mentioned frequency-analysis section 22. Here, the time series of the frequency feature parameter of the signal for recognition will be called standard pattern. 41 is the collating section, using technique, such as DP matching, performs collating with the time series (input configuration) of the frequency feature parameter of an input sound signal, and a standard pattern, and asks for the standard pattern which is best in agreement with an input configuration. 53 is ROM for characters which memorized the display pattern of a character, 53 is ROM for graphic patterns which memorized graphical display patterns, such as a picture, 61 is the display-control section, and 62 is display, such as CRT and a liquid crystal display.

[0013] The kind of standard pattern memorized by ROM33 for standard-pattern storing is determined by the capacity of ROM33 for standard-pattern storing, and the throughput of the collating section 41. If the capacity of ROM33 for standard-pattern storing is made to increase, although the kind of memorizable standard pattern will increase, time after a sound signal is inputted when the kind of standard pattern increases since the amount of operations which collating with an input configuration takes increases until it recognizes and displays becomes long. Therefore, the kind of standard pattern considers the throughput of the collating section 41, and is determined. Sound other than dozens of kinds of language, such as "good morning" required for everyday life and "being a meal" as a standard pattern in this example, and the language of some kinds, such as siren sound of a motor fire engine and an infant's cry, is memorized.

[0014] The collating section 41 is a computer in fact, and discovers the standard pattern near an input configuration using the technique of DP matching. Here, the technique of DP matching is explained briefly. Drawing 3 is drawing showing the concept of DP matching. the inside of drawing, and a1 -- a2 and a3 -- the frequency feature-parameter time series of an input, b1, b2, and b3 -- is the frequency feature-parameter time series of a standard pattern. In DP matching, distance is found, after changing a time-axis so that an input and standard-pattern frequency feature-parameter time series may be best in agreement. That is, if the path P in drawing considers as the optimal path, what totaled the difference of a and b which corresponds in each position on the P about the total position on P will be made into the distance of an input and a standard pattern. Thus, the distance about all standard patterns is found and let a standard pattern with the smallest distance be a recognition result.

[0015] Among each standard pattern, the image information corresponding to the standard pattern of a non-voice, i.e., the picture of a motor fire engine, the picture with which the infant is crying, matches with a standard pattern, and it is memorized by ROM54 for graphic patterns. The display-control section 61 performs the display corresponding to the standard pattern which is best in agreement to display 62 according to the collating result in the collating section 41. When the standard pattern to display is voice, the pattern showing the character string to display of a code and each character is read from ROM53 for characters, and is compounded, and it outputs to display 62 as a video signal. That is, the character showing the inputted language of a sound signal is displayed. In addition, it is possible to also make ROM53 for characters memorize the pattern of the character string corresponding to a standard pattern, and if it is made such, a display controller 61 can be simplified.

[0016] When the inputted sound signal is not language, the alphabetic information which shows that the sound can be heard, and a graphical display pattern is displayed. Drawing 4 is the example. (1) of drawing 4 is an example which indicates that the sound can be heard in written form when the inputted sound signal is not language, and (2) is an example which displays the

related graphical display pattern. (a) both shows the case where a siren's sound can be heard, and (b) shows the case where the baby is crying. As mentioned above, since there is a problem that a difference with the case where the others have said such language cannot be distinguished when it displays in written form, when the inputted sound signal is not language, it is desirable to add the display which shows that it decides that a graphic pattern is displayed or the others have not told (1) of drawing 4 such language further.

[0017] Although the above was the example of this invention, collating in the collating section 41 made distance what totaled the difference of a and b which corresponds in each position to the path P of the frequency feature-parameter time series of an input, and the frequency feature-parameter time series of a standard pattern about the total position on P, as shown in drawing 3 . However, it is a position corresponding to [as what was totaled about the total position is not made into distance but it is shown in drawing 5] the input configuration of the start edge and termination of Path P, respectively a1 a1 The modification of asking without fixing is also possible. In the case of the example of this drawing, it is am. an They are the start edge and termination, respectively. Thus, a part of distance of an input configuration makes it a recognition result in quest of the smallest standard pattern.

[0018] Although the sound of the object which carries out speech recognition is caught with a microphone, since two or more sound sources exist, a microphone may catch two or more sound simultaneously. In such a case, in having collated with the standard pattern as it is, it becomes difficult to specify a standard pattern in agreement. in such a case -- for example, plurality and directivity are changed, a directive microphone is formed and the input signal of each microphone is compared, in being in agreement, it is recognized as the number of sound sources being one, and performs collating processing, when the input signals of each microphone differ, it is recognized as a different sound source existing, and collating processing is performed about the sound signal from each sound source

[0019] When two or more sound signals have been recognized, it is necessary to display the sound signal recognized simultaneously. When human being's words and the siren of a motor fire engine have been recognized, while carrying out character representation of human being's words there, the picture of a motor fire engine is displayed. Moreover, when the siren of a motor fire engine and a baby's cry have been recognized, the picture of a motor fire engine and the picture over which a baby cries are displayed simultaneously. (1) of drawing 6 shows this example of a display.

[0020] Furthermore, two or more standard patterns to which distance was similar may exist as a result of collating in the collating section 41. In such a case, it is dangerous to specify the nearest standard pattern in it. Then, a display is transformed into condition of displaying two or more possible standard patterns as it is in such a case. (2) of drawing 6 is drawing showing the example of a display in such a modification, a baby's cry, the cry of a cat, or when it cannot specify completely, displays simultaneously the picture over which the baby is crying, and the picture of a cat, and shows probability to each according to a recognition result.

[0021] As mentioned above, if above equipment is used, the information to which sound other than human being's talk words is also recognized, and relates can be displayed. Although what is necessary is just to see the display screen then since it is used only when meeting a partner and talking, if talk language is recognized and displayed, there is no sound other than human being's talk words from the matte inputted when. Therefore, equipment is always made into an ON state, and although it will display when it has recognized that such sound was inputted, those who are using equipment are also considered that it cannot be recognized that there was an input of such sound. And the display screen cannot always be seen. Then, when specific sound other than human being's talk words is inputted and things have been recognized, it is made for sound and another meanses other than a display, for example, vibration etc., to report to a user. Such equipment is used for a hearing-impaired person's support equipment, and if it is made for vibration to report when the siren of a motor fire engine has been recognized, it will greatly contribute to a hearing-impaired person's safety.

[0022]

[Effect of the Invention] As explained above, according to this invention, in a voice recognition unit, it recognizes only human being's words, the sound of human being's words is recognized it not only displays in written form, but, and the information corresponding to it can be displayed now. When such equipment is used for hearing-impaired persons, a hearing-impaired person can be provided with information also including surrounding sound.

[Translation done.]

* NOTICES *

1. This document has been translated by computer.So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] A voice input means to input a sound signal (1) A feature-extraction means to extract the feature for recognizing a sound signal (2) A feature standard-pattern storage means to memorize the feature pattern of a standard sound signal (3) A collating means to collate the feature of the extracted voice input signal, and the feature pattern memorized by the aforementioned feature standard-pattern storage means (3), and to specify the standard sound signal corresponding to the inputted sound signal (4) A display pattern storage means (5) to memorize the display information corresponding to the aforementioned standard sound signal, and a display means to display the display information corresponding to a standard sound signal when a standard sound signal is specified with this collating means (4) (6) It is the voice recognition unit equipped with the above. the aforementioned feature standard-pattern storage means (3) The feature pattern (32) of non-language standard sound signals other than human being's words other than the feature pattern (31) of the standard sound signal human being's words is memorized. the aforementioned display pattern storage means (5) It is characterized by having memorized the language display information (51) which made the aforementioned standard sound signal the character as it was, and the non-language display information (52) corresponding to the aforementioned non-language standard sound signal.

[Claim 2] The aforementioned non-language display information (52) is a voice recognition unit according to claim 1 characterized by being image information and the aforementioned display means (6) being able to display image information.
[Claim 3] For the aforementioned display means (6), the image information of the aforementioned non-language display information (52) is a voice recognition unit according to claim 2 characterized by the ability to display dynamic-image information including dynamic-image information.
[Claim 4] The voice recognition unit according to claim 1 characterized by having an information means by which meanses other than voice report that information was displayed on the aforementioned display means (6).
[Claim 5] The aforementioned collating means (4) is a voice recognition unit according to claim 1 which carries out ranking ****** of the coincidence condition of the feature of the extracted voice input signal, and the feature pattern memorized by the aforementioned feature standard-pattern storage means (3), and is characterized by the aforementioned display means (6) displaying two or more display information to the standard sound signal of a predetermined coincidence condition.
[Claim 6] The aforementioned collating means (4) is a voice recognition unit according to claim 1 characterized by the aforementioned display means (6) displaying two or more display information when it is detected that two or more standard sound signals were inputted simultaneously.

[Translation done.]